STIC-ILL

From: Marx, Irene
Sent: Wednesday, August 01, 2001 12:34 PM
To: STIC-ILL
Subject: 09/582402

Please send to Irene Marx, Art Unit 1651; CM1, Room 10E05, phone 308-2922, Mail box in 11B01

Suemori et al., Semei Kogaku Kogyo Gijustsu Kenkyusho Kenkyu Hokoku (1995), 3(2), 33-36

Sparnins et al., J. Bacteriol., (1976), 127(1), 362-6

Karoum, F., Neuropsychopharmacol. Trace Amines:Exp. Clin. Aspects, 2nd (1985) 433-450. Ed. Boulton. Publisher: Humana, Clifton, N.J.

Blakley et al., Can. J. Microbiol., 1977, 23(9), 1128-1139

Blakley et al., Can. J. Microbiol., 1972, 18(8), 1247-55

Mills et al., Insect Biochem., 1971, 1(3), 264-70.

Kindl, H., Eur. J. Biochem. 1969, 7(3), 340-7

Fernandez-Canon et al., J. biol. chem., 1995, 270(36), 21199-205

Irene Marx
Art Unit 1651
CMI 10-E-05,
Mail Box 11-B-01
703-308-2922

THE JOURNAL OF BIOLOGICAL CHEMISTRY
Vol. 270, No. 36, Issue of September 8, pp. 21199-21205, 1995
Printed in U.S.A.

Molecular Characterization of a Gene Encoding a Homogentisate Dioxygenase from *Aspergillus nidulans* and Identification of Its Human and Plant Homologues*

(Received for publication, June 20, 1995, and in revised form, July 5, 1995)

José M. Fernández-Cañón and Miguel A. Peñalva‡

From the Departamento de Microbiología Molecular, Centro de Investigaciones Biológicas del CSIC, Velázquez 144, 28006 Madrid, Spain

We report here the first characterization of a gene encoding a homogentisate dioxygenase, the *Aspergillus nidulans hmgA* gene. The HmgA protein catalyzes an essential step in phenylalanine catabolism, and disruption of the gene results in accumulation of homogentisate in broths containing phenylalanine. *hmgA* putatively encodes a 448-residue polypeptide (M_r = 50,168) containing 21 histidine and 23 tyrosine residues. This polypeptide has been expressed in *Escherichia coli* as a fusion to glutathione *S*-transferase, and the affinity-purified protein has homogentisate dioxygenase activity.

***A. nidulans*, an ascomycete amenable to classical and reverse genetic analysis, is a good metabolic model to study inborn errors in human Phe catabolism. One such disease, alkaptonuria, was the first human inborn error recognized (Garrod, A. E. (1902) *Lancet* 2, 1616-1620) and results from loss of homogentisate dioxygenase. Here we take advantage of the high degree of conservation between the amino acid sequences of the fungal and higher eukaryote enzymes of this pathway to identify expressed sequence tags encoding human and plant homologues of HmgA. This is a significant advance in characterizing the genetic defect(s) of alkaptonuria and illustrates the usefulness of our fungal model.**

The physiologically versatile filamentous ascomycete *Aspergillus nidulans* is able to grow on Phe or PhAc[1] as the sole carbon source. The *A. nidulans* Phe catabolic pathway is notably similar to its human counterpart (Fernández-Cañón and Peñalva, 1995). As in humans (Fig. 1), Phe is degraded to homogentisate (2,5-dihydroxy-PhAc). The aromatic ring is then cleaved by homogentisate dioxygenase to yield, after an isomerization step, fumarylacetoacetate, which is split by fumarylacetoacetate hydrolase into fumarate and acetoacetate (see Fig. 1). *Aspergillus* can also catabolize PhAc through homogentisate after two sequential hydroxylation reactions in the aromatic ring.[2] This PhAc pathway is absent in humans.

Humans are very sensitive to defects in Phe catabolism. Loss-of-function mutations in structural genes of this pathway cause different metabolic diseases. Alkaptonuria is one such disease, resulting from loss of homogentisate dioxygenase (EC 1.13.11.15) (La Du *et al.*, 1958). This moderately disabling disease, whose main clinical features are darkening of the urine, pigmentation of cartilages, and arthritis in adults, was the first inborn error of metabolism to be described (Garrod, 1902). However, the gene encoding homogentisate dioxygenase has not been characterized from humans or any other organism (see McKusick (1994)). Therefore, definitive evidence that the disease results from a loss-of-function mutation in this gene has not yet been obtained. Type I tyrosinaemia, resulting from fumarylacetoacetate hydrolase deficiency, is a different defect in human Phe catabolism with severe consequences. Our characterization of the *fahA* gene, encoding *A. nidulans* fumarylacetoacetate hydrolase, showed 47% identity at the amino acid level with its human homologue (Fernández-Cañón and Peñalva, 1995). Loss of this enzyme results in phenylalanine toxicity and extracellular accumulation of succinylacetone, a hallmark of the disease in the urine of human patients. The similarities in the overall organization of the Phe pathway, in the amino acid sequences for at least one enzyme, and in the consequences of equivalent genetic blocks between *A. nidulans* and humans prompted us to use this fungus as a model for certain metabolic aspects of human defects in Phe catabolism. Here we use this lower eukaryote to characterize, for the first time, a gene encoding a homogentisate dioxygenase and use its deduced amino acid sequence to identify its human and plant homologues.

EXPERIMENTAL PROCEDURES

Strains, Media, and Growth Conditions—*A. nidulans biA1* was used as wild type and source of cDNA. A *biA1, methG1, argB2* strain was used as recipient for *hmgA* disruption, and a *biA1, methG1* strain was used as control in experiments involving the *hmgA*::*argB*$^+$ disrupted strain. *A. nidulans* standard media (Cove, 1966) were used for strain maintenance, growth tests, and transformation experiments. Supplements were added when necessary. For transfer experiments, *Aspergillus* liquid cultures were grown at 37 °C with vigorous shaking in an appropriately supplemented defined medium containing (in g/liter) KPO_4H_2 (13.6), $(NH_4)_2SO_4$ (2.0), $MgSO_4{\cdot}7H_2O$ (0.25), and $FeSO_4{\cdot}7H_2O$ (0.0005), with glucose at 0.3% (w/v) as carbon source. After 15-16 h of growth, mycelia were collected and transferred to fresh medium with no carbon source added. After a 1-h incubation, appropriate inducing or non-inducing carbon sources were added at the following concentrations: glucose at 1% (w/v); Phe, PhAc, all other PhAc hydroxy derivatives, and glutamate at 10 mM; Tyr at 5 mM, and potassium acetate at 30 mM. These cultures were further incubated for the times indicated in each case.

DNA Sequencing and Sequence Analysis—DNA was sequenced manually using the dideoxynucleotide chain termination method (Sanger *et al.*, 1977) with T7 DNA polymerase (Pharmacia Biotech Inc.) and ^{35}S-dATP (Amersham Corp.) on pBluescript SK$^+$ double-stranded templates. Primers were either universal or a series of oligonucleotides specific for both strands of the *hmgA* locus. Amino acid sequence similarities were found using BLAST (Altschul *et al.*, 1990) and the trans-

* The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

The nucleotide sequence(s) reported in this paper has been submitted to the GenBank™/EMBL Data Bank with accession number(s) U30797.

‡ To whom correspondence and reprint requests should be addressed. Tel.: 34-1-5611800 (ext. 4358); Fax: 34-1-5627518.

[1] The abbreviations used are: PhAc, phenylacetate; EST, expressed sequence tag; GST, glutathione *S*-transferase; IPTG, isopropyl-1-thio-β-D-galactopyranoside; HPLC, high performance liquid chromatography; ORF, open reading frame.

[2] J. M. Fernández-Cañón and M. A. Peñalva, unpublished results.

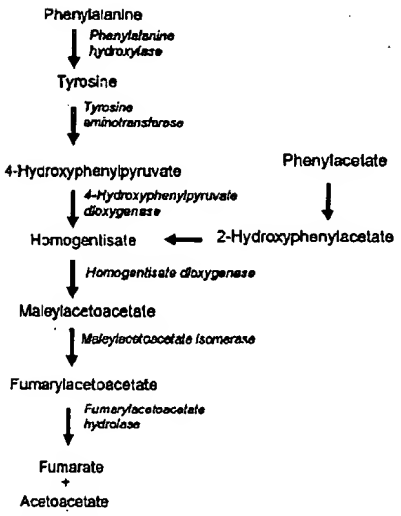

FIG. 1. **The Phe/PhAc pathway in *A. nidulans*.** The enzymatic reactions involved in Phe catabolism are the same as in humans. Humans do not use the PhAc pathway.

lation of nucleotide sequences in GenBank and EMBL data bases. Alignments were further improved by eye. Only amino acid identities were considered.

Transformation Techniques and Gene Disruption—A. nidulans was transformed following the method of Tilburn *et al.* (1983). *hmgA* was disrupted by transformation with an *argB*$^+$-containing plasmid carrying a 1.05-kilobase *ClaI-EcoRI* fragment internal to the *hmgA* transcriptional unit. This fragment starts within the first intron splitting the coding region (at residue 23) and ends at residue 347. Thus, of the two incomplete genes generated after targeting this plasmid to *hmgA*, one lacks the 101 C-terminal codons while the second lacks the promoter region, the first exon, and part of the first intron.

cDNA Cloning, Libraries, and Transcript Analysis—hmgA cDNA clones were isolated from a λgt10 cDNA library made with transcripts isolated from PhAc-induced cells. This library was screened with "plus" and "minus" cDNA probes as described (Fernández-Cañón and Peñalva, 1995). The "plus" probe was a subtracted cDNA probe obtained following Zimmermann *et al.* (1980) and Sargent (1987). Mycelia for RNA isolation were obtained after transfer to indicated carbon sources as described above and further incubation for 1 h at 37 °C. RNA isolation and Northern analysis were essentially as in Espeso and Peñalva (1992). The *hmgA* probe for Northern analysis and isolation of genomic clones (from a λEMBL4 library) was a cDNA clone insert containing the complete *hmgA* ORF.

Homogentisate Dioxygenase Assays—The enzyme was assayed spectrophotometrically following formation of maleylacetoacetate at 330 nm (Edwards and Knox, 1955). Mycelia pregrown in glucose were transferred to 10 mM PhAc defined medium (strongly inducing conditions) and incubated for 1 h at 37 °C. Lysis was achieved by sonication in 100 mM potassium phosphate buffer, pH 7.0. Extracts were clarified by centrifugation (13,000 × *g*, 10 min, 4 °C) and used in enzyme assays, which contained, in a final volume of 1 ml, 100 mM potassium phosphate buffer, pH 7.0, 2 mM ascorbate, 50 μM $FeSO_4$, 250 μM homogentisate, and 50 μl of extract (50–75 μg of protein). Recombinant GST::HmgA(9–448) enzyme expressed in *Escherichia coli* was assayed by the same method.

Construction of GST Fusion Proteins and Expression in E. coli—A cDNA insert containing the complete ORF (including the stop codon) was obtained from a cDNA clone as a *NotI* fragment. Ends were filled in with Klenow polymerase, and the fragment was cloned in the filled-in *AvaI* site of pGEX-2T (Pharmacia). One of the two possible orientations resulted in a plasmid driving expression in *E. coli* of a GST::HmgA(9–448) fusion protein (predicted molecular mass, 76 kDa). This plasmid was named pGEX::HMGA. The opposite orientation (plasmid pGEX::AGMH) encodes an aberrant fusion protein terminating shortly after the end of the N-terminal GST moiety. A third pGEX derivative (pGEX::+2) was constructed in which the *hmgA* ORF was fused to GST in the correct translational orientation but in the incorrect reading frame, resulting in premature termination 13 residues after the end of GST. *E. coli* strains carrying each of the above plasmids were cultured at 28 °C in LB broth with ampicillin (100 μg·ml^{-1}) until an A^{660} = 0.6 was reached. At this point 100 μM IPTG was added and cultures incubated for 3 additional hours at 28 °C. Cells were chilled in ice, concentrated by centrifugation and resuspension in 15 mM potassium phosphate buffer, pH 7.0, 150 mM NaCl, and lysed by sonication. Extracts were clarified by centrifugation (13,000 × *g*, 30 min, 4 °C) and either used directly in enzyme assays or purified (in the case of bacteria carrying pGEX::HMGA) by glutathione-Sepharose (Pharmacia) affinity chromatography, essentially as described by the manufacturer. Elution buffer was exchanged with 15 mM potassium phosphate, pH 7.0, 150 mM NaCl using a Pharmacia PD-10 column before using the purified preparation in enzyme assays.

HPLC Analysis of Culture Filtrates—Mycelia of the *hmgA::argB*$^+$ and of the *hmgA*$^+$ (otherwise isogenic) control strain were pregrown in defined medium with glucose, transferred to flasks containing this medium with 25 mM Phe as sole carbon source, and incubated for 8 additional hours at 37 °C. Culture supernatants were filtered through 0.45-μm pore size filters. 20-μl samples of 1:5 dilutions were analyzed with an HPLC apparatus equipped with a variable wavelength UV-visible detector and a Nucleosil 300–5 C-18 column (4 mm (inner diameter) × 250 mm), which was coupled to a 11 × 4-mm precolumn of the same support. The mobile phase was 50 mM sodium phosphate buffer (pH 6.5)/methanol (80:20, v/v) at a flow rate of 0.5 ml·min^{-1}.

RESULTS

Cloning and Molecular Characterization of the hmgA Gene—We have used a subtraction procedure to isolate cDNAs for *A. nidulans* genes whose transcription is induced by utilization of PhAc as sole carbon source (Fernández-Cañón and Peñalva, 1995). A class of these cDNAs was present in the λgt10 library with a relative abundance of 2%. Nucleotide sequencing of overlapping cDNA clones and comparison with the corresponding genomic sequences revealed the presence of a long ORF interrupted by three short introns, 48, 54, and 49 nucleotides, respectively (Fig. 2). This ORF can encode a 448-residue polypeptide, whose amino acid sequence showed no significant similarity to sequences with an assigned function deposited in SwissProt and PIR data bases or those obtained by conceptual translation of GenBank/EMBL data base entries (but see also below).

Northern blot hybridization experiments (Fig. 3) showed that this gene encodes an ~1.7-kilobase message whose transcription pattern conforms to that expected for a gene of Phe/PhAc catabolism. The gene was strongly induced by PhAc, Phe, and certain monohydroxy derivatives of PhAc and weakly induced by PhAc dihydroxy derivatives (on which the fungus grows very poorly). Transcription was undetectable under carbon starvation conditions or in the presence of either glucose or either of two gluconeogenic carbon sources (acetate and glutamate). This pattern strongly suggests that expression of this gene is specifically induced by PhAc and Phe or by a common catabolite. Glucose only slightly reduces induction by PhAc. This transcription pattern is identical to that of the *fahA* gene (Fernández-Cañón and Peñalva, 1995). Indeed, Southern blot hybridization experiments using DNA from purified phages from a λEMBL4 genomic library showed that this gene is closely linked to *fahA*, encoding *A. nidulans* fumarylacetoacetate hydrolase. Nucleotide sequencing demonstrated that both genes were transcribed divergently, their corresponding ORFs being separated by an intergenic region of 415 base pair(s),[2] presumably containing common elements controlling transcription of both genes. Clustering of genes belonging to the same metabolic pathway is not unusual in filamentous ascomycetes. This, together with the above transcription data, strongly suggested that we had isolated a previously uncharacterized gene of the Phe/PhAc catabolic pathway. The deduced molecular mass of its encoded polypeptide is 50,168 Da, which is very similar to that of 49 kDa estimated by SDS-polyacrylamide gel electrophoresis for purified murine homogentisate dioxygenase (Schmidt *et al.*, 1995). Thus, we suspected that this gene might encode a fungal homogentisate dioxygenase.

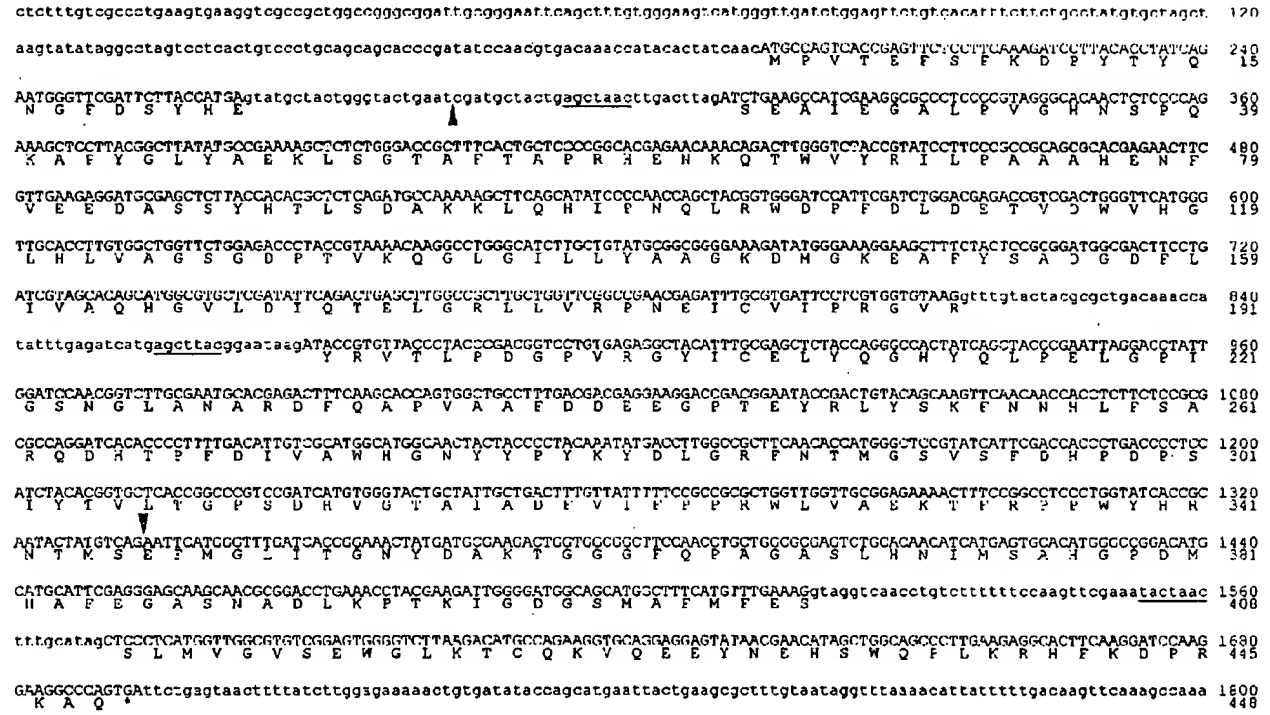

FIG. 2. **Nucleotide sequence of the *hmgA* gene and amino acid sequence of its putative translation product.** The nucleotide sequence shown was determined in both strands by sequencing a genomic and several overlapping cDNA clones covering the complete open reading frame. Coding and non-coding sequences are indicated in *uppercase* and *lowercase letters*, respectively. Introns, all flanked by consensus donor and acceptor sequences, were determined by comparison of genomic and cDNA sequences. Putative lariat boxes, as determined by comparison to the consensus sequence described by Parker *et al.* (1987) in yeast, are *underlined*. *Arrows* indicate restriction enzyme cuts used to construct the disruption plasmid (see below).

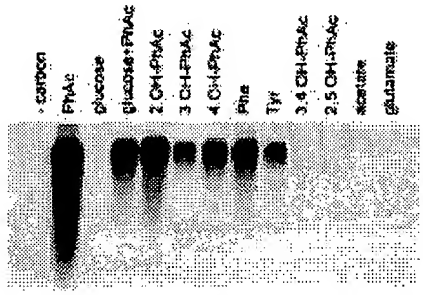

FIG. 3. **Transcription of *hmgA* is induced by PhAc or Phe.** Northern blot hybridization of total RNA isolated from mycelia transferred to carbon sources indicated (− carbon indicates no carbon source added; see "Experimental Procedures" for concentrations of carbon sources). The probe was a ^{32}P-labeled cDNA containing the complete *hmgA* ORF. Loading in each *lane* was equal, as determined by hybridization with an actin probe (not shown).

We describe below genetic and biochemical evidence demonstrating that this is indeed the case, the *Aspergillus* homogentisate dioxygenase gene being designated *hmgA*.

Disruption of the hmgA Gene Prevents Growth on Phe and Results in Secretion of a Red Pigment—In order to determine the loss-of-function phenotype of *hmgA*, a plasmid containing a genomic DNA fragment internal to the ORF was targeted to the *hmgA* locus by homologous recombination (Fig. 4*A*). The 5′-end of this fragment is located within the first intron (thereby lacking the promoter, the initiation codon, and the first exon), and its 3′-end is at codon 347 (thus lacking the 101 C-terminal codons). Homokaryotic transformants in which integration took place by single-copy homologous recombination in the resident *hmgA* gene were selected and verified by Southern analysis. Two independent transformants showing the expected hybridization pattern were phenotypically identical, and only one was studied further. This strain, carrying the *hmgA*::*argB*$^+$ disruption, grew normally on lactose or on mixtures of lactose and Phe or PhAc but was unable to grow on either Phe or PhAc as sole carbon source (Fig. 4*B*). In addition, *hmgA* disruption resulted in secretion of a red pigment on growth media containing Phe or PhAc (Fig. 4*B*). This pigment is visible even on Phe or PhAc plates, on which there was only residual mycelial growth of the mutant strain (due to utilization of other carbon sources present in the agar). The *hmgA*::*argB*$^+$ gene disruption is recessive in diploids both for growth on Phe/PhAc plates and for secretion of the red pigment. Thus, *hmgA* integrity is required for Phe/PhAc catabolism, and its loss-of-function results in external accumulation of a red pigment.

Disruption of hmgA Results in Secretion of Homogentisate due to Complete Absence of Homogentisate Dioxygenase—Human alkaptonuria (resulting from loss of homogentisate dioxygenase) causes ochronosis of connective tissues due to the accumulation of a dark pigment, which is an oxidation product of homogentisate. Similarly, the red pigment produced by the *hmgA*::*argB*$^+$ strain turned dark brown after prolonged incubation times on culture plates or after a much shorter period in shaken (more aerated) liquid cultures. This suggested that this strain might secrete homogentisate. HPLC analysis of filtrates of cultures pregrown in glucose and transferred to Phe minimal medium (Fig. 5*A*) confirmed that, in contrast to the *hmgA*$^+$ strain, the *hmgA*::*argB*$^+$ strain secreted homogentisate. This was strongly indicative of a block in homogentisate dioxygenase. Indeed, while substantial homogentisate dioxygenase activity was induced in *hmgA*$^+$ mycelia upon transfer to Phe, no

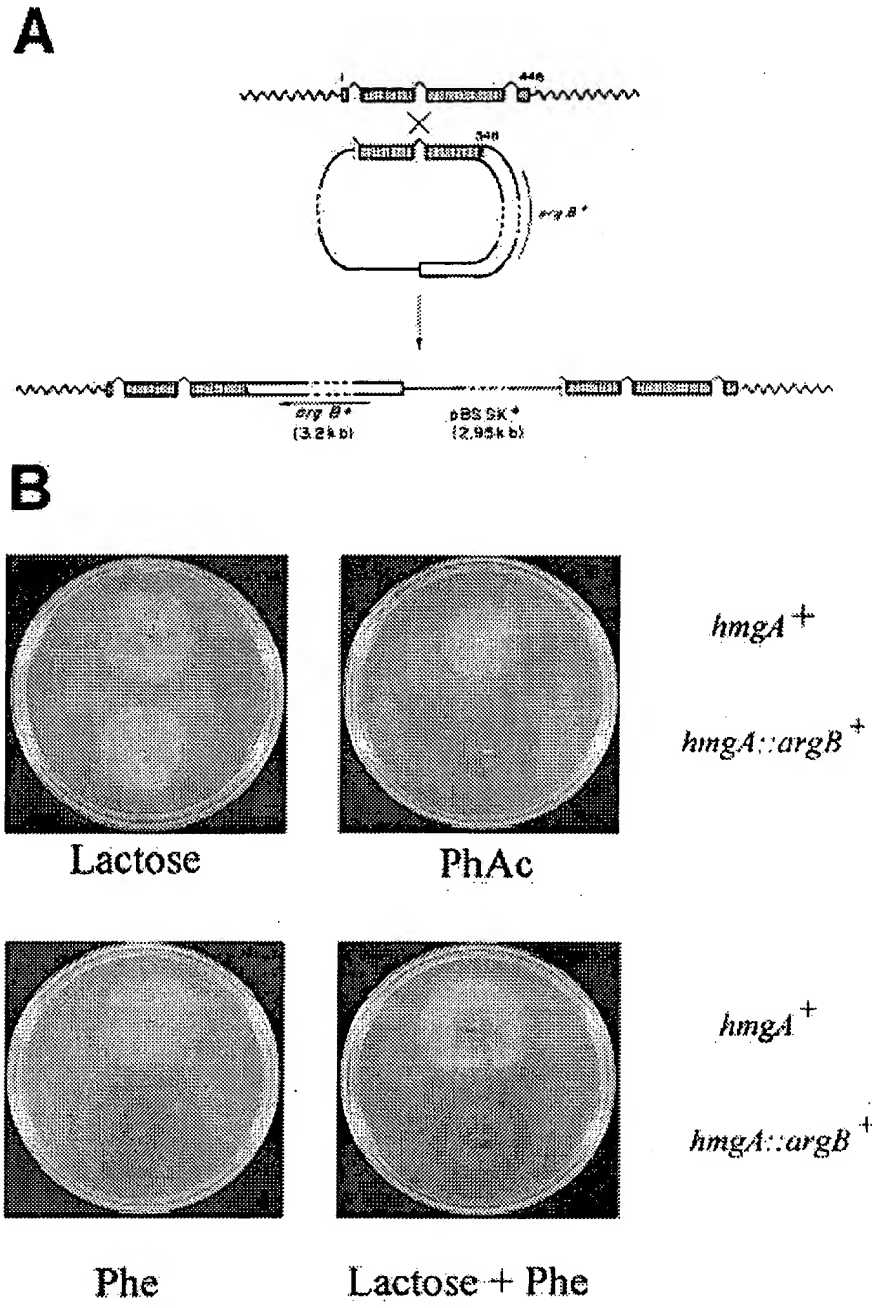

FIG. 4. **Disruption of the *hmgA* gene and growth tests.** *Panel A*, a plasmid containing a internal fragment of *hmgA* was used in a transformation experiment to disrupt the *hmgA* gene, as shown in the scheme (see "Experimental Procedures" for details). This integration event results in two incomplete copies of *hmgA*, one lacking the 101 C-terminal codons and the other lacking the promoter region, the first exon, and part of the first intron. Two different transformants having identical phenotypes had this integration event, as established by Southern blot analysis using *hmgA* and *argB* probes (data not shown). *Open boxes* denote the *argB*$^+$ gene; *lined boxes* denote the *hmgA* coding region (with three introns indicated); a *wavy line* indicates *A. nidulans* coding sequences; and a *straight line* indicates vector DNA. *Numbers* above the *hmgA* coding region indicate codon position. *Panel B*, growth phenotypes of the disrupted strain. Conidia of the *hmgA*$^+$ and of the *hmgA*::*argB*$^+$ strains were inoculated on minimal medium plates with the indicated carbon sources (lactose at 0.05% (w/v), PhAc at 10 mM, and Phe at 25 mM). Plates were incubated for 3 days at 37 °C before being photographed.

such activity was detectable in mycelial extracts from the mutant strain (Fig. 5*B*). We conclude that loss-of-function of *hmgA* results in absence of homogentisate dioxygenase and consequent accumulation of homogentisate.

hmgA Encodes a Protein with Homogentisate Dioxygenase Activity—To establish that *hmgA* encodes a homogentisate dioxygenase, we expressed its protein product tagged with GST (Smith and Johnson, 1988) in *E. coli*. A *Not*I cDNA fragment filled in with Klenow enzyme was subcloned in the filled-in *Ava*I site of pGEX-2T. Recombinant plasmids in either orientation were recovered. The correct orientation (plasmid pGEX::HMGA) resulted in an in-frame fusion between the N-terminal GST coding region and codons 9–449 of the *hmgA* coding region (including the stop codon). The resulting fusion protein (predicted molecular mass, 76 kDa) was designated GST::HmgA (9–448). In contrast, in the reverse orientation (plasmid pGEX::AGMH), the GST reading frame is truncated by several in-frame stop codons present in the non-coding strand of the *hmgA* gene. As a second control, a third plasmid was constructed in which the *hmgA* coding region was fused to GST in the correct orientation but in an incorrect reading frame (pGEX::+2). Expression of the fusion genes in bacterial strains respectively carrying each of these three plasmids was induced with IPTG and homogentisate dioxygenase assayed in protein extracts. No activity was detectable in extracts of strains containing either the pGEX::AGMH or pGEX::+2 plasmids. In contrast, high levels of homogentisate dioxygenase were detected in the extract corresponding to the GST fusion in the correct orientation and reading frame (Fig. 6*A*). A 76-kDa band (*i.e.* the expected molecular mass for the GST::HmgA (9–448) fusion protein) was induced by IPTG in bacteria carrying pGEX::HMGA (Fig. 6*B*) but was absent in bacteria car-

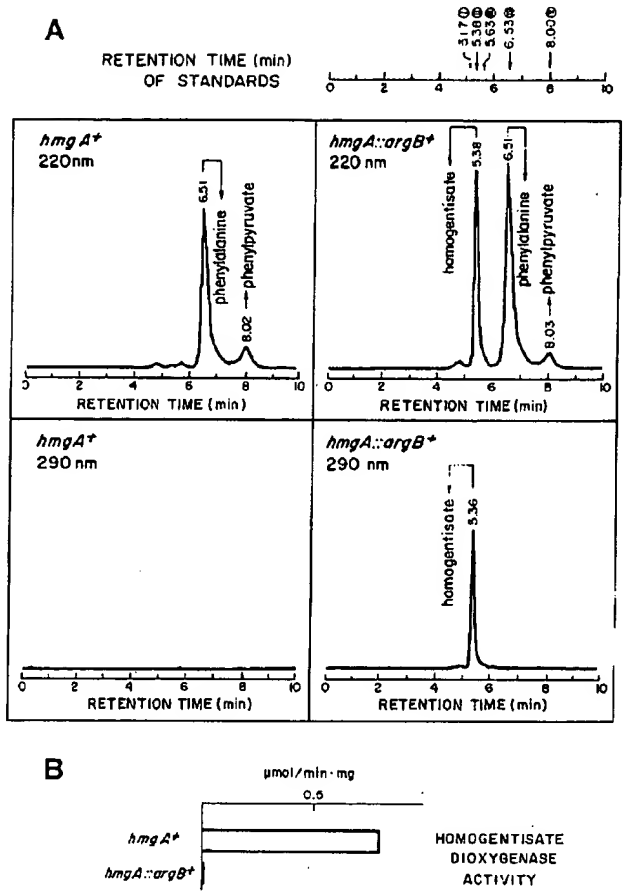

FIG. 5. **Disruption of *hmgA* results in secretion of homogentisate and absence of homogentisate dioxygenase.** *Panel A*, HPLC analysis of culture filtrates of the *hmgA*$^+$ and the *hmgA::argB*$^+$ strains after transfer to defined medium with 25 mM Phe as sole carbon source. The detector was set at the wavelengths indicated. Homogentisate has a characteristic absorption maximum at 290 nm. Standards were as follows: *I*, 3,4-dihydroxyPhAc; *II*, homogentisate; *III*, 2,5-dihydroxybenzoate; *IV*, Phe; *V*, phenylpyruvate. *Panel B*, homogentisate dioxygenase activity in mycelial extracts from the wild type and the *hmgA::argB*$^+$ strains.

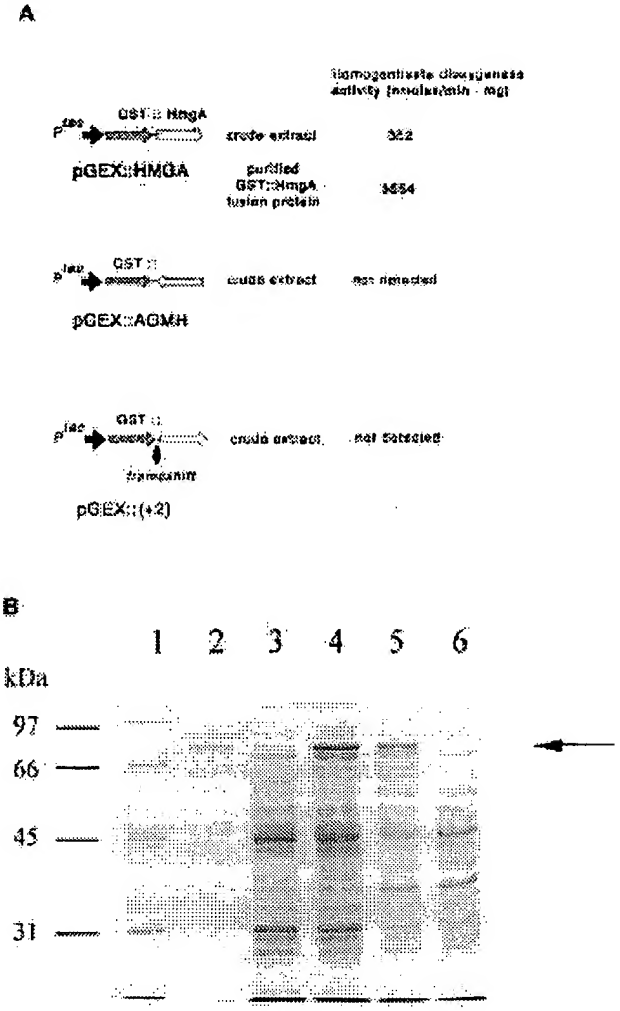

FIG. 6. **Expression in *E. coli* and purification of a GST::HmgA(9–448) fusion protein with homogentisate dioxygenase activity.** *Panel A*, assays of homogentisate dioxygenase activity in crude and purified extracts from bacteria carrying the plasmids indicated. The scheme shows relevant features of the GST fusion genes. P^{tac} indicates the promoter of pGEX2T. *Panel B*, SDS-polyacrylamide electrophoresis of the GST::HmgA(9–448) protein fusion expressed in *E. coli*. Samples were as follows: *lane 1*, size standards (Bio-Rad); *lane 2*, purified fraction after elution with glutathione; *lane 3*, crude extract after passage through the glutathione-Sepharose column; *lane 4*, crude bacterial extract; *lanes 5* and *6*, cells lysed in sample buffer, 3 h after induction (*lane 5*) or without IPTG induction (*lane 6*). Note the ~76-kDa band induced by IPTG, retained by the affinity column, and eluted with glutathione (indicated by an *arrow*).

rying either of the control plasmids (data not shown). To prove that induction of homogentisate dioxygenase resulted from expression of the GST::HmgA (9–448) fusion protein, it was affinity-purified from crude extracts. This purification resulted in a nearly homogeneous preparation having 10 times higher specific activity than crude extracts (see Fig. 6, *A* and *B*). We conclude that *hmgA* encodes a homogentisate dioxygenase.

We noticed a significant loss in the activity of the GST::HmgA (9–448) fusion protein after purification by glutathione-Sepharose chromatography. The specific activity of our purified fusion protein (Fig. 6*A*) is 7.5 times higher than that reported for murine homogentisate dioxygenase (Schmidt *et al.*, 1995). These figures represent a minimal estimation for the specific activity of the *Aspergillus* enzyme, as our protein has been produced in a heterologous host and we cannot exclude a negative effect of the N-terminal GST moiety on enzyme activity. Our purified GST::HmgA (9–448) fusion protein showed no activity when the reaction was carried out in the absence of Fe^{2+} ions, as described for purified murine homogentisate dioxygenase (Schmidt *et al.*, 1995). In contrast, the enzyme was fully active in the absence of ascorbate.

Identification of Human and Plant Homologues of hmgA—A search of the conceptual translation products of GenBank and EMBL data bases with the HmgA amino acid sequence revealed no similarity to any polypeptide encoded by an entry with an assigned function. However, ESTs encoding deduced polypeptides with significant sequence identity to HmgA were identified (Fig. 7). Four of these are human ESTs, of which three were from either liver or mixed liver/spleen libraries. Two of them (apart from probable sequence ambiguities) appear to be identical, their deduced amino acid sequence showing 50% identity to a 93-residue region in the C-terminal half of *A. nidulans* HmgA. The deduced amino acid sequence of the third showed 42% identity to an 89-residue N-terminal region of HmgA (Fig. 7). A fourth EST, putatively encoding a polypeptide showing similarity to this N-terminal region, was from pancreatic islets. This amino acid sequence differed from that of the liver clone in only three different positions.

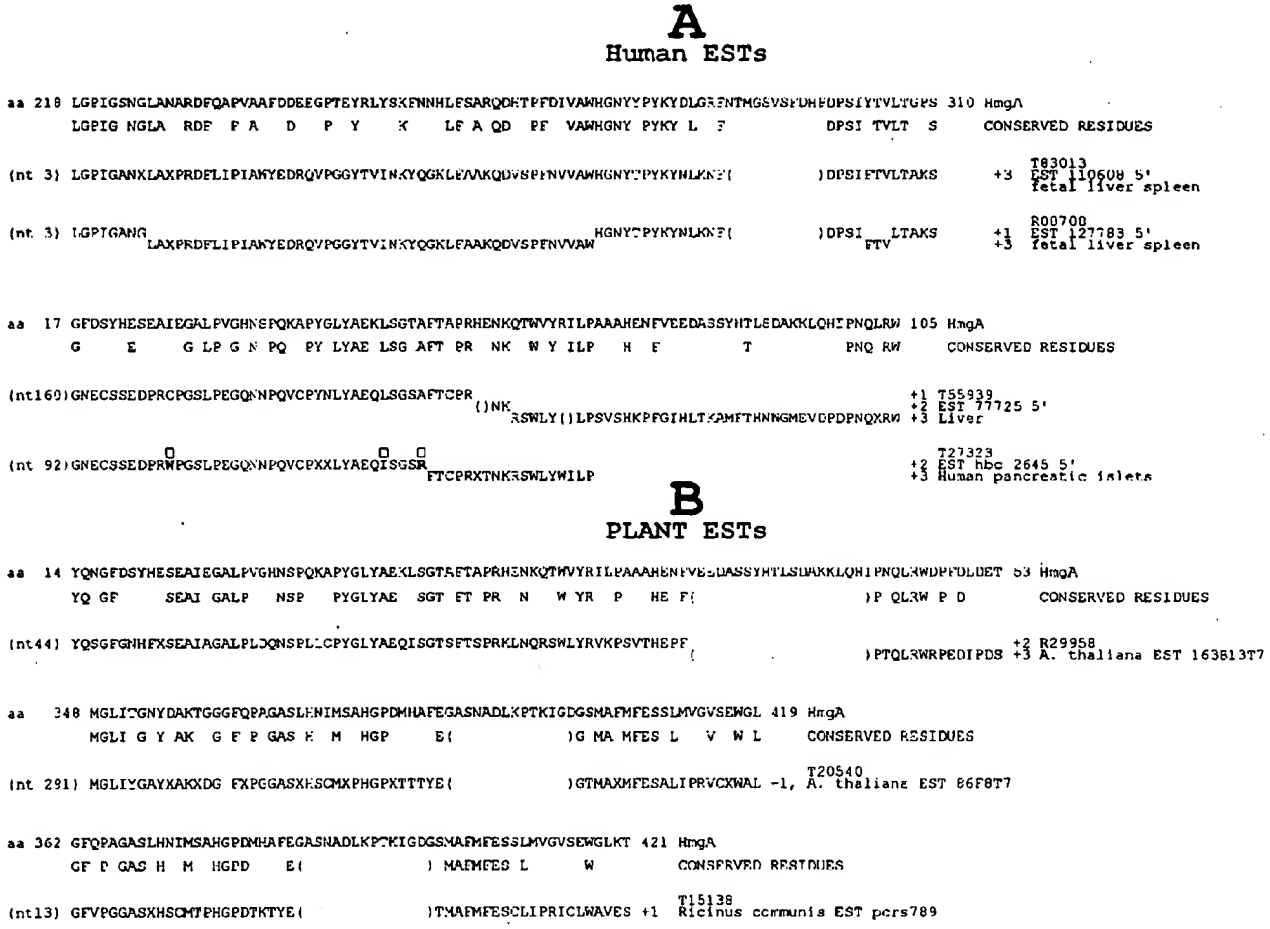

FIG. 7. **Identification of human and plant homologues of the *hmgA* gene.** Sequence comparison between HmgA and deduced polypeptides encoded by human and plant ESTs. *A*, human 5′-ESTs. Identical residues are shown below the indicated regions of HmgA. Deduced sequences encoded by ESTs were grouped in two pairs, each corresponding to a different region of HmgA. It should be noted that sequences codified by T55939 and T27323 differ in three positions (residues in *boldface* (indicated by *boxes*); see also text). Conserved residues are not indicated for clarity. Residues marked as *X* represent ambiguities in the nucleotide sequence of the EST. *Numbers* on the *right* and *left ends* of HmgA sequences denote the positions of the amino acid residues in the protein. *Numbers* (*in parentheses*) on the *left side* of each EST polypeptide refer to the nucleotide position of the first translated codon in its corresponding DNA sequence. In several cases, frameshifting allows alignments with significant sequence conservation, possibly reflecting sequencing errors in ESTs. Therefore, deduced sequences corresponding to different reading frames (indicated on the *right*) of the same 5′-EST are shown when necessary. *Parentheses* indicate regions of deduced sequence with no evident similarity in any of the three EST reading frames. Also indicated on the *right* are the GenBank accession numbers and the source of cDNA. The liver spleen library is a mixed library. *B*, plant 5′-ESTs. Sequence comparisons were shown as in *A*, indicating the organism from which the library was constructed.

The remaining three ESTs were from plants, two being from *Arabidopsis*. Their deduced polypeptides showed 44 and 43% identity to non-overlapping HmgA regions close to the N- and C-terminal regions, respectively. The polypeptide encoded by the third plant EST (from *Ricinus communis*) has 36% identity to the above near C-terminal HmgA region.

DISCUSSION

We describe here the first characterization in any organism of a gene encoding an homogentisate dioxygenase, the *A. nidulans hmgA* gene. Homogentisate dioxygenase activity is strongly induced in mycelia by Phe or PhAc, and *hmgA* cDNA clones were easily isolated using a subtracted cDNA probe from a library enriched in cDNAs for PhAc-induced transcripts. *hmgA* is essential for growth on Phe (or PhAc) as sole carbon source. This supports the *A. nidulans* Phe (PhAc) catabolic pathway as shown in Fig. 1. Disruption of the gene results in complete absence of homogentisate dioxygenase activity. Therefore, the disruption created here is almost certainly a null allele. Due to this defect, this strain, when supplied with Phe or PhAc, secretes homogentisate, which is readily oxidized to yield a reddish pigment, eventually turning dark brown. The complete absence of enzyme activity in the disruption strain together with the absence of cross-hybridizing bands in genomic Southern blots strongly indicates that *A. nidulans* contains a single gene encoding homogentisate dioxygenase.

Mammalian homogentisate dioxygenases contain weakly bound ferrous ions that are required for activity (see Schmidt *et al.* (1995) and references therein). A 70% reduction in activity was also observed[2] when crude *Aspergillus* extracts were assayed in the absence of Fe^{2+} ions, and the activity of the purified GST::HmgA (9–448) fusion protein showed an absolute requirement for these ions. The deduced HmgA polypeptide contains 21 His and 23 Tyr residues. Some of these residues might be involved in binding iron, as demonstrated for protocatechuate 3,4-dioxygenase (Ohlendorf *et al.*, 1988).

The deduced HmgA sequence was used to identify human ESTs potentially encoding a homologue(s) of the fungal gene. The high similarity detected at the amino acid level establishes

that human genes corresponding to these ESTs are indeed *hmgA* homologues. The EST-encoded amino acid sequences are classified in two groups, each corresponding to a different region of HmgA. Therefore, our results do not establish the existence of a single human gene encoding homogentisate dioxygenase.

Three of these ESTs were isolated from liver or liver/spleen cDNA libraries. Alkaptonuria results from loss of homogentisate dioxygenase, as demonstrated in both liver and kidney extracts. Our preliminary identification of cDNAs for homogentisate dioxygenase, nearly a century after alkaptonuria was recognized by Garrod (1902) as an inborn error of metabolism, represents a significant advance in the characterization of the human gene and further illustrates the validity of our fungal metabolic model for disorders in human Phe metabolism (Fernández-Cañón and Peñalva, 1995). Definitive evidence that alkaptonuria results from loss-of-function mutation(s) in the homogentisate dioxygenase gene will require mapping of this gene to chromosome 3q2 (the location for alkaptonuria (Pollak *et al.*, 1993; Janocha *et al.*, 1994)) and identification of mutations in patients with the syndrome. Finally, availability of the human gene will allow examination of tissue-specific expression. One of the ESTs identified here was isolated from pancreatic islet cDNA. Homogentisate dioxygenase activity in this tissue has not been reported previously.

We have previously reported a class of mutations (*suAfah*) suppressing the effects of a complete fumarylacetoacetate hydrolase deficiency in *A. nidulans* (Fernández-Cañón and Peñalva, 1995). The phenotype of these suppressor mutations is indistinguishable of that caused by *hmgA*::*argB*$^+$ mutation (*i.e.* loss of homogentisate dioxygenase and secretion of homogentisate), and they do not complement the disruption in diploids. These *suAfah* mutations are therefore *hmgA*$^-$ alleles (Fernández-Cañón and Peñalva, 1995). This suppression led us to propose that alkaptonuria would prevent the lethal effects of human type I tyrosinaemia by blocking the pathway upstream of fumarylacetoacetate hydrolase. Identification of mammalian homologues of *hmgA* reported here will facilitate testing this hypothesis in animal models.

Acknowledgments—We thank E. Reoyo for technical assistance, R. Uña and A. Hurtado for artwork, and H. N. Arst, Jr., C. Scazzocchio, S. Rodríguez de Córdoba, and J. L. García for discussions and critical reading of the manuscript.

REFERENCES

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990) *J. Mol. Biol.* **215**, 403–410
Cove, D. J. (1966) *Biochim. Biophys. Acta* **113**, 51–56
Edwards, S. W., and Knox, W. E. (1955) *Methods Enzymol.* **2**, 292–295
Espeso, E. A., and Peñalva, M. A. (1992) *Mol. Microbiol.* **6**, 1457–1465
Fernández-Cañón, J. M., and Peñalva M. A. (1995) *Proc. Natl. Acad. Sci. U. S. A.*, in press
Garrod, A. E. (1902) *Lancet* **2**, 1616–1620
Janocha, J., Wolz, W., Srsen, S., Srsnova, K., Montagutelli, X., Guénet, J. L., Grimm, T., Kress, W., and Müller, C. R. (1994) *Genomics* **19**, 5–8
La Du, B. N., Zannoni, V. G., Laster, L., and Seegmiller, J. E. (1958) *J. Biol. Chem.* **230**, 251–260
McKusick, V. A. (1994) *Genomics* **19**, 3–4
Ohlendorf, D. H., Lipscomb, J. D., and Weber, P. C. (1988) *Nature* **336**, 403–405
Parker, R., Siciliano, P. G., and Guthrie, C. (1987) *Cell* **49**, 229–239
Pollak, M. R., Chou, Y. W., Cerda, J. J., Steinmann, B., La Du, B. N., Seidman, J. G., and Seidman, C. E. (1993) *Nat. Genet.* **5**, 201–204
Sanger, F., Nicklen, S., and Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U. S. A.* **74**, 5463–5467
Sargent, T. D. (1987) *Methods Enzymol.* **152**, 423–432
Schmidt, S. R., Müller, C. R., and Kress, W. (1995) *Eur. J. Biochem.* **228**, 425–430
Smith, D. B., and Johnson, K. S. (1988) *Gene (Amst.)* **67**, 31–40
Tilburn, J., Scazzocchio, C., Taylor, G. G., Zabicky-Zissman, J. H., Lockington, R. A., and Davies, R. W. (1983) *Gene (Amst.)* **26**, 205–221
Zimmermann, C. R., Orr, W. C., Leclerc, R. F., Barnard, E. C., and Timberlake, W. E. (1980) *Cell* **21**, 709–715